Methods for Screening for Transdominant Effector Peptides and RNA Molecules

Inventors: Garry P. Nolan and S. Michael Rothenberg
Attorney: Richard Aron Osman, Ph.D., Reg. #36,627

INTRODUCTION

Technical Field

The technical field of this invention is methods for screening for transdominant effector peptides and RNA molecules selected inside living cells from randomized pools.

Background

Signaling pathways in cells often begin with an effector stimulus that leads to a phenotypically describable change in cellular physiology. Despite the key role intracellular signaling pathways play in pathogenesis, in most cases, little is understood about a signaling pathway other than the initial stimulus and the ultimate cellular response.

Historically, signal transduction has been analyzed by biochemistry or genetics. The biochemical approach dissects a pathway in a "stepping-stone" fashion: find a molecule that acts at, or is involved in, one end of the pathway, isolate assayable quantities and then try to determine the next molecule in the pathway, either upstream or downstream of the isolated one. The genetic approach is classically a "shot in the dark": induce or derive mutants in a signaling pathway and map the locus by genetic crosses or complement the mutation with a cDNA library. Limitations of biochemical approaches include a reliance on a significant amount of pre-existing knowledge about the constituents under study and the need to carry such studies out *in vitro*, post-mortem. Limitations of purely genetic approaches include the need to first derive and then characterize before proceeding with identifying and cloning the gene.

Screening molecular libraries of chemical compounds for drugs that regulate signaling

systems has led to important discoveries of great clinical significance. Cyclosporin A (CsA) and FK506, for examples, were selected in standard pharmaceutical screens for inhibition of T-cell activation. It is noteworthy that while these two drugs bind completely different cellular proteins - cyclophilin and FK506 binding protein (FKBP), respectively, the effect of either drug is virtually the same - profound and specific suppression of T-cell activation, phenotypically observable in T cells as inhibition of mRNA production dependent on transcription factors such as NF-AT and NF-κB. Libraries of small peptides have also been successfully screened in assays for bioactivity. The literature is replete with examples of small peptides capable of modulating a wide variety of signaling pathways. For example, a peptide derived from the HIV-1 envelope protein has been shown to block the action of cellular calmodulin.

A major limitation of conventional *in vitro* screens is delivery. While only minute amounts of an agent may be necessary to modulate a particular cellular response, delivering such an amount to the requisite subcellular location necessitates exposing the target cell or system to relatively massive concentrations of the agent. The effect of such concentrations may well mask or preclude the targeted response.

The present invention provides methods and compositions to create, effectively introduce into cells and screen compounds that affect a signaling pathway. Llittle or no knowledge of the pathway is required, other than a presumed signaling event and an observable physiologic change in the target cell. The disclosed methods are conceptually distinct from prior library search methods in that it is an *in vivo* stratagem for accessing intracellular signaling mechanisms. The invention also provides for the isolation of the constituents of the pathway, the tools to characterize the pathway, and lead compounds for pharmaceutical development.

<u>Relevant Literature</u>

Mann et al. (1983) Cell 33, 153-159, Pear et al. (1993) Proc. Natl. Acad. Sci. USA 90(18):8392-6 and copending US patent application Serial No. 08/023,909 describe the BOSC and BING retroviral systems useful as delivery vectors for the disclosed methods.

Scott and Craig (1994) Current Opinion in Biotechnology 5:40-48 review random peptide libraries. Hupp et al. (1995) describe small peptides which activate the latent sequence-specific DNA binding function of p53. Palzkill et al. (1994) report the selection of functional signal

cleavage sites from a library of random sequences introduced into TEM-1 -lactamase.

SUMMARY OF THE INVENTION

The invention provides methods and compositions for screening for transdominant intracellularly bioactive agents such as pharmaceuticals. The invention accesses molecules or targets within living cells and provides for the direct selection of those bioactive agents with desired phenotypic effects. The general methods involve steps: expressing a molecular library of randomized nucleic acids as a plurality of isolated corresponding randomized expression products in a plurality of cells, each of the nucleic acids comprising a different nucleotide sequence, screening for a cell of the plurality of cells exhibiting a changed physiology in response to the presence in the cell of a transdominant expression product of the corresponding expression products; and detecting and isolating the cell and/or transdominant expression product,

In a particular embodiment, the expressing step comprises translating the nucleic acids and/or corresponding transcripts, and each of the nucleic acids encodes a peptide comprising a different amino acid sequence. The nucleic acids may be joined to sequences encoding polypeptide backbones of artificial design capable of intracellularly presenting randomized peptides as structured domains. The methods may also involve introducing the library into the cells, such as through the use of retroviral vectors, and particularly suitable vectors are also disclosed.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Creation of a library of random peptides in a retrovirus DNA construct by PCR.

Figure 2. Creation of a library of random peptides in a retrovirus DNA construct by primed DNA synthesis.

Figure 3. Presentation constructs for localizing presentation structures to specific cellular locales.

Figure 4. Schematic of a retroviral construct.

DETAILED DESCRIPTION OF THE INVENTION

The subject methods generally involve introducing a molecular library of randomized nucleic acids into a population of cells. The introduced nucleic acids are randomized and expressed in the cells as a library of isolated randomized expression products, which may be nucleic acids, such as message, antisense RNA, ribozyme components, etc., or peptides. The library should provide a sufficiently structurally diverse population of randomized expression products to effect a probabilistically sufficient range of cellular responses to provide one or more cells exhibiting a desired response. Generally at least 10^6, preferably at least 10^7 more preferably at least 10^8 and most preferably at least 10^9 different expression products are simultaneously analyzed in the subject methods. Preferred methods maximize library size and diversity.

The introduced nucleic acids and resultant expression products are randomized, meaning that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. The library may be fully random or biased, e.g. in nucleotide/residue frequency generally or per position. For example, a biased library may encode peptides for interactions with known classes of molecules, such as SH-3 domain proteins, as defined by peptides containing XXXPPXPXX (where X=randomized residues). In other embodiments, the nucleotides or residues are randomized within a defined class, e.g. of hydrophobic amino acids, of purines, etc. In any event, where the ultimate expression product is a nucleic acid, at least 10, preferably at least 12, more preferably at least 15, most preferably at least 21 nucleotide positions need to be randomized; more if the randomization is less than perfect. Similarly, at least 5, preferably at least 6, more preferably at least 7 amino acid positions need to be randomized; again, more if the randomization is less than perfect.

An important aspect of the invention is the functional and structural isolation of the randomized expression products. This is important to facilitate subsequent drug development and optimization. Generally, isolation is effected by providing free (not covalently coupled) expression product, though in some situations, the expression product may be coupled to a functional group or fusion partner, preferably a heterologous (to the host cell) or synthetic (not native to any cell) functional group or fusion partner. Exemplary groups or partners include signal sequences capable of constitutively localizing the expression product to a predetermined subcellular locale such as the Golgi, endoplasmic reticulum, nucleoli, nucleus, nuclear membrane,

mitochondria, chloroplast, secretory vesicles, lysosome, etc.; binding sequences capable of binding the expression product to a predetermined protein while retaining bioactivity of the expression product; sequences signaling selective degradation, of itself or co-bound proteins; secretory and membrane-anchoring signals; etc.

It may also be desirable to provide a partner which conformationally restricts the randomized expression product to more specifically define the number of structural conformations available to the cell. For example, such a partner may be a synthetic presentation structure: an artificial polypeptide capable of intracellularly presenting a randomized peptide as a conformation-restricted domain. Generally such presentation structures comprise a first portion joined to the N-terminal end of the randomized peptide, and a second portion joined to the C-terminal end of the peptide. Preferred presentation structures maximize accessibility to the peptide by presenting it on an exterior loop, for example of coiled-coils, (Myszka, D.G., and Chaiken, I.M. Design and characterization of an intramolecular antiparallel coiled coil peptide. Biochemistry. 1994. 33:2362-2372). To increase the functional isolation of the randomized expression product, the presentation structures are selected or designed to have minimal biologically active as expressed in the target cell. In addition, the presentation structures may be modified, randomized, and/or matured to alter the presentation orientation of the randomized expression product. For example, determinants at the base of the loop may be modified to slightly modify the internal loop peptide tertiary structure, while maintaining the absolute amino acid identity. Other presentation structures include zinc-finger domains, loops on beta-sheet turns and coiled-coil stem structures in which non-critical residues are randomized; loop structures held together by cysteine bridges, cyclic peptides, etc.

Following expression of the library in the targeted cells, cells exhibiting a changed physiology in response to the presence in such cells of a transdominant expression product are detected and isolated. A transdominant expression product has an effect that is not in cis, i.e., a trans event as defined in genetic terms or biochemical terms. A transdominant effect is a distinguishable effect by a molecular entity (i.e., the encoded peptide or RNA) upon some separate and distinguishable target; that is, not an effect upon the encoded entity itself. As such, transdominant effects include many well-known effects by pharmacologic agents upon target

molecules or pathways in cells or physiologic systems; for instance, the β-lactam antibiotics have a transdominant effect upon peptidoglycan synthesis in bacterial cells by binding to penicillin binding proteins and disrupting their functions. An exemplary transdominant effect by a peptide is the ability to inhibit NF-κB signaling by binding to IκB-α at a region critical for its function, such that in the presence of sufficient amounts of the peptide (or molecular entity), the signaling pathways that normally lead to the activation of NF-κB through phosphorylation and/or degradation of IκB-α are inhibited from acting at IκB-α because of the binding of the peptide or molecular entity. In another instance, signaling pathways that are normally activated to secrete IgE are inhibited in the presence of peptide. Or, signaling pathways in adipose tissue cells, normally quiescent, are activated to metabolize fat. Or, in the presence of a peptide, intracellular mechanisms for the replication of certain viruses, such as HIV-1, or Herpes viridae family members, or Respiratory SyncytiaVirus, for example, are inhibited.

A transdominant effect upon a protein or molecular pathway is clearly distinguishable from randomization, change, or mutation of a sequence within a protein or molecule of known or unknown function to enhance or diminish a biochemical ability that protein or molecule already manifests. For instance, a protein that enzymatically cleaves β-lactam antibiotics, a β-lactamase, could be enhanced or diminished in its activity by mutating sequences internal to its structure that enhance or diminish the ability of this enzyme to act upon and cleave β-lactam antibiotics. This would be called a cis mutation to the protein. The effect of this protein upon β-lactam antibiotics is an activity the protein already manifests, to a distinguishable degree. Similarly, a mutation in the leader sequence that enhanced the export of this protein to the extracellular spaces wherein it might encounter β-lactam molecules more readily, or a mutation within the sequence that enhance the stability of the protein, would be termed cis mutations in the protein. For comparison, a transdominant effector of this protein would include an agent, independent of the β-lactamase, that bound to the β-lactamase in such a way that it enhanced or diminished the function of the β-lactamase by virtue of its binding to β-lactamase.

In general, cis-effects are effects within molecules wherein elements that are interacting are covalently joined to each other although these elements might individually manifest themselves as separable domains. Trans-effects (transdominant in that under some cellular conditions the desired

effect is manifested) are those effects between distinct molecular entities, such that molecular entity A, not covalently linked to molecular entity B, binds to or otherwise has an effect upon the activities of entity B. As such, most known pharmacological agents are transdominant effectors.

A wide variety of phenotypic changes may be targeted, such as changes in the amount or distribution of cellular differentiation markers, e.g. T/B-cell activation, cell growth or death, membrane potentials, etc. Similarly, a wide variety of techniques may be used to isolate cells exhibiting changed physiology in response to the presence of a library expression product, such as FACS, lysis selection using complement, cell cloning, etc. Where the randomized expression product is expressed on the extracellular surface of the host cell or is secreted by the host cell into the extracellular medium, the methods may involve screening for a change in the physiology of a "responder cell" in the vicinity of the host cell, exhibiting a changed physiology in response to the presence of the surface-expressed or secreted expression product. Alternatively, the methods may involve screening for a "target cell" in the vicinity of the host cell, that specifically binds the surface-expressed or secreted expression product.

A wide variety of cells may be used depending on the nature of the targeted cellular pathway, so long as the cell is suitably transfectable with the library and may be monitored for the targeted physiological changes. The method of introduction is largely dictated by the targeted cell type. Exemplary methods include $CaPO_4$ precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc. As many pharmaceutically important screens require human or model mammalian cell targets, retroviral vectors capable of transfecting such targets are frequently employed. A particularly well-suited retroviral transfection system is described in Mann et al.(1983) Cell 33, 153-159, Pear et al. (1993) Proc. Natl. Acad. Sci. USA 90(18):8392-6 and copending US patent application Serial No. 08/023,909.

In one embodiment of the invention, the library is generated in a retrovirus DNA construct backbone. Preferred techniques for constructing such libraries are described in more detail below. Libraries with up to 10^9 unique sequences can be readily generated in such DNA backbones. After generation of the library in DNA, and its isolation from bacteria, it is converted into infectious virus. Conversion into infectious virus requires the delivery of the library DNA to a retrovirus packaging system. Such systems can include BOSC23, PhiNX-eco, PhiNX-ampho,

293T + gag-pol and retrovirus envelope, PA317, and other extant packaging systems. Methods for high-efficiency packaging of retroviral libraries (cDNA) are described in Kitamura et al. (1995) *Efficient screening of retroviral cDNA libraries.* PNAS 92: 9146-9150. Libraries containing up to 10^7 retrovirus particles per ml were derived in this system. Scaling up these methods conveniently yields 10^9 individual and unique library components in a 100 ml reaction. The library of retroviruses is then used to infect the target cells by any of a number of different protocols, e.g. see Kitamura et al *supra* for details on some infection protocols

Depending upon how the particular resident expression product(s) effect the target molecule function, the physiology of the cell will be affected in differing ways. Since extremely large libraries of different molecular peptide or RNA shapes can be delivered to cells via retroviral transduction, this embodiment of the invention provides a high frequency for the isolation of cells affected in a predetermined manner. For instance, peptides that inhibit T cell activation processes are readily isolated using a population of Jurkat cells, e.g. Jurkat-NFAT-dipA, engineered to express diphtheria toxin under the control of Nuclear Factor of Activated T cells (NFAT) regulatory elements. Unproductively transduced cells are killed when NFAT is activated through the T cell receptor and other signaling mechanisms specific to T cells (Serafini AT et al. (1995). *Isolation of mutant T lymphocytes with defects in capacitative calcium entry.* Immunity, 3(2):239-50). This is because NFAT activation leads to transcription and translation of diphtheria toxin A chain in these cells, which efficiently kills them. However, productive mutations in the NFAT activation pathway are rescued. This approach has already been used to isolate mutants in T cell signaling pathways. Peptide or compound expression libraries are also used to mimic mutation of the signaling pathway leading to activation of NFAT, i.e. peptides which interfere with signaling of the NFAT pathway by binding to and inhibiting the function of a required signaling protein or factor can block programmed cell death.

The following experimental section/examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

<u>Retroviral Presentation Construct for Peptide Expression.</u>

A series of retroviral constructs have been designed for expression of randomized and biased peptides within target cell populations. The peptide is expressed from a retroviral promoter. The translation unit has several important components. Glycine following the initiator methionine at the amino terminus stabilizes the peptide and enhances cytoplasmic half-life, according to Varshavsky's N-End Rule. In some constructs, a nine amino acid flu epitope tag has been incorporated to permit co-precipitation of the rare peptide and any molecule to which it has affinity, by using monoclonal antibodies to the epitope. Glycines are encoded before and after the random/biased expression product encoding regions to provide some molecular flexibility. Two carboxyl-terminal prolines are encoded to confer stability to carboxypeptidase.

For construction of a large library two primers were made (schematized in Figure 1). The first, designated the random peptide primer, consists of 1) a complementary region for vector priming, 2) the regions mentioned above, and 3) a random or biased expression product region, here presented as a 30 base sequence encoding a peptide of length 10 amino acids. In addition, we have inserted a stop codon in all three reading frames in case of minor deletions or insertions in the random region. The design of the primer ensures a glycine/proline termination in most reading frames. The second primer is downstream in the vector and primes a region of the plasmid that contains a unique Not I site. These primers are used to create a library of fragments, each containing a different nucleotide sequence that each potentially encodes a different peptide. These families of fragments are ligated to vector fragments containing puromycin selection sequence, a 3' LTR , and a bacterial origin of replication. The ligation products are then electroporated into E. coli and DNA is prepared from the resulting library. Using this technique we have constructed independent random libraries with up to 2 x 10^8 unique inserts. Sequencing of multiple individual inserts demonstrates they have the structure as defined by Primer 1, and the peptides encoded are random. Such libraries thus made contain subsets of the total 10^{13} predicted peptides.

<u>Generation of Retroviral Peptide Libraries.</u>

A scheme for generating a peptide library in the pBabe Puro vector is shown in Figure 2. Primers for PCR were synthesized, purified and deprotected according to standard protocols. Primer 1, complementary to polylinker sequences in the pBabe Puro retroviral construct, has the

sequence 5' GCT TAG CAA GAT CTC TAC GGT GGA CCK NNK NNK NNK NNK NNK NNK NNK NNK NNK NNC CCC ACT CCC ATG GTC CTA CGT ACC ACC ACA CTG GG 3'. N represents any of the four bases; K is limited to G or T. Primer 2 has the sequence 5' GCT TAG CAA GAT CTG TGT GTC AGT TAG GGT GTG G 3' and is complementary to sequences within the pUC18 origin of replication. PCR was carried out for 8 rounds using primer 1, primer 2, pBabe Puro as template, and a mixture of Taq DNA Polymerase (Promega) and Deep Vent DNA Polymerase (New England Biolabs) in a ratio of 128 Taq: 1 Deep Vent as described in Barnes (1994) Proc. Natl. Acad. Sci. USA, 91, pp. 2216-2220. The amplified PCR product was purified, digested with restriction enzymes Bgl II and Not I (Promega), purified again and ligated with the corresponding Bam HI-Not I fragment of pBabe Puro. After transformation the resulting library contained ~$2x10^8$ clones, greater than 80% of which contained inserts.

<u>pMSCV-PC and pBabeMN-PC retroviral construct libraries</u>:

Oligonucleotides were synthesized and purified according to standard protocols. The "library" oligonucleotides have the sequence 5' CTG GAG AAC CAG GAC CAT GGG C $(NNK)_{10}$ GGG CCC CCT TAA ACC ATT AAA T 3' or 5' CTG GAG AAC CAG GAC CAT GGG CNN KNN KNN KCC TCC CNN KCC TNN KNN KGG GCC CCC TTA AAC CAT TAA AT 3'. A third oligonucleotide ("constant"), complementary to the 3' ends of the library oligonucleotides, has the sequence 5' TCA TGC ATC CAA TTT AAT GGT TTA AG 3'. As shown in Fig. 3, each library oligonucleotide is annealed to the constant oligonucleotide, converted to double stranded DNA with Sequenase (United States Biochemical) or Klenow (Promega), digested with restriction enzyme Bst XI (New England Biolabs), and purified and ligated with the appropriate Bst XI-digested retroviral construct. Transformation efficiencies are ~ 2×10^8 clones per microgram of ligated DNA, greater than 90% of which contain an insert. A representative retrovirus is shown in Fig. 4; see also, retroviral nucleotide sequence below:

<u>Retroviral vector with presentation construct.</u>

TGAAAGACCCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGAAA
ATACATAACTGAGAATAGAGAAGTTCAGATCAAGGTTAGGAACAGAGAGACAGCAGAATATGGGC
CAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAG
ATGCGGTCCCGCCCTCAGCAGTTTCTAGAGAACCATCAGATGTTTCCAGGGTGCCCCAAGGACCT
GAAAATGACCCTGTGCCTTATTTGAACTAACCAATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCT

```
TCTGCTCCCCGAGCTCAATAAAAGAGCCCACAACCCCTCACTCGGCGCGCCAGTCCTCCGATAGA
CTGCGTCGCCCGGGTACCCGTATTCCCAATAAAGCCTCTTGCTGTTTGCATCCGAATCGTGGACT
CGCTGATCCTTGGGAGGGTCTCCTCAGATTGATTGACTGCCCACCTCGGGGGTCTTTCATTTGGA
GGTTCCACCGAGATTTGGAGACCCCTGCCTAGGGACCACCGACCCCCCCGCCGGGAGGTAAGCTG
GCCAGCGGTCGTTTCGTGTCTGTCTCTGTCTTTGTGCGTGTTTGTGCCGGCATCTAATGTTTGCG
CCTGCGTCTGTACTAGTTAGCTAACTAGCTCTGTATCTGGCGGACCCGTGGTGGAACTGACGAGT
TCTGAACACCCGGCCGCAACCCTGGGAGACGTCCCAGGGACTTTGGGGGCCGTTTTTGTGGCCCG
ACCTGAGGAAGGGAGTCGATGTGGAATCCGACCCCGTCAGGATATGTGGTTCTGGTAGGAGACGA
GAACCTAAAACAGTTCCCGCCTCCGTCTGAATTTTTGCTTTCGGTTTGGAACCGAAGCCGCGCGT
CTTGTCTGCTGCAGCGCTGCAGCATCGTTCTGTGTTCTCTCTGTCTGACTGTGTTTCTGTATTTG
TCTGAAAATTAGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCACTGGAAAGATGT
CGAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGACGTTGGGTTACCTTCTGCTCTG
CAGAATGGCCAACCTTTAACGTCGGATGGCCGCGAGACGGCACCTTTAACCGAGACCTCATCACC
CAGGTTAAGATCAAGGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGGTCCCCTACATCGT
GACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCTTTGTACACCCTAAGCCTC
CGCCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAACCTCCTCGTTCGACCCCGCCTCGA
TCCTCCCTTTATCCAGCCCTCACTCCTTCTCTAGGCGCCGGAATTccaggaccatgggcGGGCCC
CCTTAAAccattaaattggtaaaataaagGATCCGTCGACCTGCAGCCAAGCTTATCGATAAAAT
AAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGGAATGAAAGACCCCACCTGTAGGTTTGGCAAG
CTAGCTTAAGTAACGCCATTTTGCAAGGCATGGAAAATACATAACTGAGAATAGAGAAGTTCAGA
TCAAGGTTAGGAACAGAGAGACAGCAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCC
TGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCCGCCCTCAGCAGTTTCTAGA
GAACCATCAGATGTTTCCAGGGTGCCCCAAGGACCTGAAATGACCCTGTGCCTTATTTGAACTA
ACCAATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCC
ACAACCCCTCACTCGGCGCGCCAGTCCTCCGATAGACTGCGTCGCCCGGGTACCCGTGTATCCAA
TAAACCCTCTTGCAGTTGCATCCGACTTGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGT
GATTGACTACCCGTCAGCGGGGGTCTTTCATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA
ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGT
GCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC
GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCA
GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTG
```

AGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC
TCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC
GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCT
GTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTT
CAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTT
ATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG
AGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTG
CTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG
TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATC
CTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTC
ATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAAT
CTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCT
CAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATA
CGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTAT
CCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGT
TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC
ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGG
TTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTT
ATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGA
GTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA
TACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCG
GGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACC
CAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAA
ATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAA
TATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCA
TTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTC
GGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGC
GGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGC
TTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCAC

AGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGA
AGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGC
GATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCACG
CTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCG
CCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCC
ACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGT
GATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTC
CGGCGTAGAG

<u>Peptide Library Infection of a Factor-dependent Line and Outgrowth of an Apoptosis-Resistant Line.</u>

The Baf/3 cell line is an IL-3 dependent cell that undergoes rapid apoptosis in the absence of IL-3. Thus it makes an attractive cell line for dominant effector peptides. Cells expressing a peptide that inhibits apoptosis are readily selected against the background of dying cells. We chose this cell line as a model for demonstrating peptide selection.

A retroviral library containing 5 x 10^6 independent peptide inserts was transfected into BOSC23 cells and converted into retrovirus with an approximate titer of 5 x 10^5 per ml. Twelve ml of viral supernatant was used to infect 6 x 10^6 Baf/3cells (2 ml per infection of 1 x 10^6 cells in 6 independent infections). Cells were grown for 3 days after infection in the presence of IL-3 to allow retroviral integration and peptide expression. After three days IL-3 was withdrawn and the cells allowed to grow for two weeks. After two weeks, one well of six had outgrowth of cells that survive in the absence of IL-3, indicating the presence of an apoptosis-inhibiting peptide. Peptides derived in this manner may effect the IL-3 independence by positive dominancy (i.e. mimic or circumvent the positive regulatory role of IL-3) or by inhibition (i.e. prevent the apoptosis process upon IL-3 withdrawal).

<u>Subcellular Targeting:</u>

In some embodiments of the invention, expression products are localized to, or preferentially concentrated in, different subcellular compartments within cells, e.g. by using appropriate addition of addressins to a peptide presentation construct, see, Figure 3. Addressins are available for a wide variety of subcellular locales including the nucleus, Golgi, mitochondria, plasma membranes, endoplasmic reticulum, secretory granules, secreted, cell surface (extracellular

domain with random), cell surface (intracellular domain random), etc. For example, many proteins whose functions require entry into the cell nucleus include nuclear localization signal (NLS) sequences: generally short, positively charged (basic) domains that serve to direct the entire protein in which they occur to the cell's nucleus. Numerous NLS amino acid sequences have been reported including single basic NLS's such as that of the SV40 (monkey virus) large T antigen (Pro Lys Lys Lys Arg Lys Val), Kalderon (1984), et al., Cell, 39:499-509, and double basic NLS's exemplified by that of the Xenopus (African clawed toad) protein, nucleoplasmin (Ala Val Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys LysLys Leu Asp), Dingwall, et al., Cell, 30:449-458, 1982 and Dingwall, et al., J.Cell Biol., 107:641-849; 1988). Numerous localization studies have demonstrated that NLSs incorporated in synthetic peptides or grafted onto reporter proteins not normally targeted to the cell nucleus cause these peptides and reporter proteins to be concentrated in the nucleus. See, for example, Dingwall, and Laskey, Ann, Rev. Cell Biol., 2:367-390, 1986; Bonnerot, etal., Proc, Natl. Acad, Sci, USA, 84:6795-6799, 1987; Galileo, et al., Proc. Natl. Acad. Sci. USA, 87:458-462, 1990.

<u>Secreted Peptide Structure.</u>

In some embodiments of the invention, it is desired to generate a peptide capable of binding to the surface of, or affecting the physiology of, a target cell that is other than the host cell, e.g the cell infected with the retrovirus. In this case the peptide library is configured with a secretion signal sequence at its amino terminus, usually clipped off after secretion, such that the peptide is ultimately secreted into the extracellular space. Secretory signal sequences and their transferability to unrelated proteins are well known, e.g. Silhavy et al.(1985) Microbiol Rev 49, 398-418. In a preferred approach, a fusion product is configured to contain, in series, secretion signal peptide-presentation structure-randomized expression product region-presentation structure, see Figure 3. In this manner, target cells grown in the vicinity of cells caused to express the library of peptides, are bathed in secreted peptide. Target cells exhibiting a physiological change in response to the presence of a peptide, e.g. by the peptide binding to a surface receptor or by being internalized and binding to intracellular targets, and the secreting cells are localized by any of a variety of selection schemes and the peptide causing the effect determined. Exemplary effects include variously that of a designer cytokine (i.e., a stem cell factor capable of causing

hematopoietic stem cells to divide and maintain their totipotential), a factor causing cancer cells to undergo spontaneous apoptosis, a factor that binds to the cell surface of target cells and labels them specifically, etc.

Surface-expressed peptide structure.

In certain embodiments of the invention, it is desirable to localize the randomized peptides to the surface of a cell. The invention provides methods for presenting the randomized expression product extracellularly or in the cytoplasmic space; see Fig 3. For extracellular presentation, a modification of the secreted structure above is made; specifically, a membrane anchoring region is provided at the carboxyl terminus of the peptide presentation structure. The anchor can be a transmembrane domain, such as that found in a variety of surface expressed molecules of the Ig family (e.g. CD4, CD8, IgM, T cell receptor), or a lipid anchoring region such as found in Thy-1. Other anchoring domains exist and are known to those who practice in the art. The randomized expression product region is expressed on the cell surface and presented to the extracellular space, such that it can bind to other surface molecules (affecting their function) or molecules present in the extracellular medium. The binding of such molecules could confer function on the cells expressing a peptide that binds the molecule. The cytoplasmic region could be neutral or could contain a domain that, when the extracellular randomized expression product region is bound, confers a function on the cells (activation of a kinase, phosphatase, binding of other cellular components to effect function). Similarly, the randomized expression product-containing region could be contained within a cytoplasmic region, and the transmembrane region and extracellular region remain constant or have a defined function.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.